

Machine Learning WORKSHEET –1

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans- Least Square Error

2. Which of the following statement is true about outliers in linear regression?

Ans- Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

Ans- Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans- Correlation

5. Which of the following is the reason for over fitting condition?

Ans- Low bias and high variance

6. If output involves label then that model is called as:

Ans- All of the above

7. Lasso and Ridge regression techniques belong to _____?

Ans- Regularization

8. To overcome with imbalance dataset which technique can be used?

Ans- SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans- TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Ans- True

11. Pick the feature extraction from below:

- A) Construction bag of words from a email**
- B) Apply PCA to project high dimensional data**
- C) Removing stop words**
- D) Forward selection**

Ans- Apply PCA to project high dimensional data

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

- A) We don't have to choose the learning rate.**
- B) It becomes slow when number of features is very large.**
- C) We need to iterate.**
- D) It does not make use of dependent variable.**

Ans - It becomes slow when number of features is very large.

b) It becomes slow when number of features is very large.

D) It does not make use of dependent variable.

13. Explain the term regularization?

Ans- Regularization consists of different techniques and methods used to address the issue of over-fitting by reducing the generalization error without affecting the training error much. Choosing overly complex models for the training data points can often lead to overfitting.

Types of Regularization

- 1. L1/L2 regularization
- 2. Drop out
- 3. Early Stopping

14. Which particular algorithms are used for regularization?

Ans - Lasso Regression :

- Lasso regression is regularization technique to reduce the complexity of the model. It stands for least absolute and selection operator.
- It is similar to the ridge regression except that the penalty term contains only the absolute weight instead of a square of a weights.

- Since it takes absolute values, hence, it can shrink the slope to 0, whereas ridge regression can only shrink to 0.
- It is also called as L1 regularization. Some of the features in this technique are complete neglected for model evaluation.
- Hence, the lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

15. Explain the term error present in linear regression equation?

Ans – An error term in statistics is a value which represents how observed data differs from actual population data. It can also be a variable which represents how a given statical models differ from reality. The error term is often written Σ .

It is often said that the error term in a regression equation represents the effect of the variables that were omitted from the equation. This is unsatisfactory, even in simple contexts, as a following discussion should indicate. Suppose subjects are IID, and all variables are jointly normal with expectation 0. suppose the explanatory variables have variance 1. The explanatory variables may be correlated amongst themselves, but any of them have a non-singular p -dimensional distribution. The parameters α_i are real. Let

(1) $Y_i = \sum_{j=1}^p \alpha_j X_{ij}$ for each $i=1, 2, \dots$, consider the regression model

(2) $Y_i = \sum_{j=1}^p \alpha_j X_{ij} + \epsilon_i$ where (

(3) $\epsilon_i = \sum_{j=p+1}^{\infty} \alpha_j X_{ij}$

The α_j are identifiable. If the X_{ij} are independent for $j=1, 2, \dots$, the standard assumption hold and $l(p)$ does indeed represent the effect on Y_i of the omitted variables $\{X_{ij} : j=1, 2, \dots, P\}$